NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN ISOPRENOID SYNTHESIS

INTRODUCTION

This application claims the benefit of the filing date of the provisional Application U.S. Serial Number 60/129,899, filed April 15, 1999, and the provisional Application, U.S. Serial Number 60/146,461, filed July 30, 1999.

TECHNICAL FIELD

The present invention is directed to nucleic acid and amino acid sequences and constructs, and methods related thereto.

BACKGROUND

Isoprenoids are ubiquitous compounds found in all living organisms. Plants synthesize a diverse array of greater than 22,000 isoprenoids (Connolly and Hill (1992) *Dictionary of Terpenoids*, Chapman and Hall, New York, NY). In plants, isoprenoids play essential roles in particular cell functions such as production of sterols, contributing to eukaryotic membrane architecture, acyclic polyprenoids found in the side chain of ubiquinone and plastoquinone, growth regulators like abscisic acid, gibberellins, brassinosteroids or the photosynthetic pigments chlorophylls and carotenoids. Although the physiological role of other plant isoprenoids is less evident, like that of the vast array of secondary metabolites, some are known to play key roles mediating the adaptative responses to different environmental challenges. In spite of the remarkable diversity of structure and function, all isoprenoids originate from a single metabolic precursor, isopentenyl diphosphate (IPP) (Wright, (1961) *Annu. Rev. Biochem.* 20:525-548; and Spurgeon and Porter, (1981) in Biosynthesis of Isoprenoid Compounds., Porter and Spurgeon eds (John Wiley, New York) Vol. 1, pp1-46).

A number of unique and interconnected biochemical pathways derived from the isoprenoid pathway leading to secondary metabolites, including tocopherols, exist in chloroplasts of higher plants. Tocopherols not only perform vital functions in plants, but are also important from mammalian nutritional perspectives. In plastids, tocopherols account for up to 40% of the total quinone pool.

Tocopherols and tocotrienols (unsaturated tocopherol derivatives) are well known antioxidants, and play an important role in protecting cells from free radical damage, and in the prevention of many diseases, including cardiac disease, cancer, cataracts, retinopathy, Alzheimer's disease, andneurodegeneration, and have been shown to have beneficial effects on symptoms of arthritis, and in anti-aging. Vitamin E is used in chicken feed for improving the shelf life, appearance, flavor, and oxidative stability of meat, and to transfer tocols from feed to eggs. Vitamin E has been shown to be essential for normal reproduction, improves overall performance, and enhances immunocompetence in livestock animals. Vitamin E supplement in animal feed also imparts oxidative stability to milk products.

The demand for natural tocopherols as supplements has been steadily growing at a rate of 10-20% for the past three years. At present, the demand exceeds the supply for natural tocopherols, which are known to be more biopotent than racemic mixtures of synthetically produced tocopherols. Naturally occurring tocopherols are all *d*-stereomers, whereas synthetic α-tocopherol is a mixture of eight *d,l*-α-tocopherol isomers, only one of which (12.5%) is identical to the natural *d*-α-tocopherol. Natural *d*-α-tocopherol has the highest vitamin E activity (1.49 IU/mg) when compared to other natural tocopherols or tocotrienols. The synthetic α-tocopherol has a vitamin E activity of 1.1 IU/mg. In 1995, the worldwide market for raw refined tocopherols was $1020 million; synthetic materials comprised 85-88% of the market, the remaining 12-15% being natural materials. The best sources of natural tocopherols and tocotrienols are vegetable oils and grain products. Currently, most of the natural Vitamin E is produced from γ-tocopherol derived from soy oil processing, which is subsequently converted to α-tocopherol by chemical modification (α-tocopherol exhibits the greatest biological activity).

Methods of enhancing the levels of tocopherols and tocotrienols in plants, especially levels of the more desirable compounds that can be used directly, without chemical modification, would be useful to the art as such molecules exhibit better functionality and biovailability.

In addition, methods for the increased production of other isoprenoid derived compounds in a host plant cell is desirable. Furthermore, methods for the production of particular isoprenoid compounds in a host plant cell is also needed.

SUMMARY OF THE INVENTION

The present invention is directed to D-1-deoxyxylulose 5-phosphate reductoisomerase (dxr), and in particular to dxr polynucleotides and polypeptides. The polynucleotides and polypeptides of the present invention include those derived from eukaryotic sources.

Thus, one aspect of the present invention relates to isolated polynucleotide sequences encoding D-1-deoxyxylulose 5-phosphate reductoisomerase proteins. In particular, isolated nucleic acid sequences encoding dxr proteins from plant sources are provided.

Another aspect of the present invention relates to oligonucleotides which include partial or complete dxr encoding sequences.

It is also an aspect of the present invention to provide recombinant DNA constructs which can be used for transcription or transcription and translation (expression) of dxr. In particular, constructs are provided which are capable of transcription or transcription and translation in host cells.

In another aspect of the present invention, methods are provided for production of dxr in a host cell or progeny thereof. In particular, host cells are transformed or transfected with a DNA construct which can be used for transcription or transcription and translation of dxr. The recombinant cells which contain dxr are also part of the present invention.

In a further aspect, the present invention relates to methods of using polynucleotide and polypeptide sequences to modify the isoprenoid content of host cells, particularly in host plant cells. Plant cells having such a modified isoprenoid content are also contemplated herein.

The modified plants, seeds and oils obtained by the expression of the dxr are also considered part of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides an amino acid alignment between the Arabidopsis dxr sequence and the E coli dxr sequence

Figure 2 provides a schematic diagram of the isoprenoid pathway, both the mevalonate and non-mevalonate pathways.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides, *inter alia*, compositions and methods for altering (for example, increasing and decreasing) the isoprenoid levels and/or modulating their ratios in host cells. In particular, the present invention provides polynucleotides, polypeptides, and methods of use thereof for the modulation of isoprenoid content in host plant cells.

Isoprenoids are derived from a 5- carbon building block, isopentenyl diphosphate (IPP), which is the universal isoprene unit and common isoprenoid precursor. Isoprenoids comprise a structurally diverse group of compounds that can be classified into two classes; primary and secondary metabolites (Chappell (1995) *Annu Rev. Plant Physiol. Plant Mol. Biol.* 46:521-547). Primary metabolites comprise those isoprenoids which are necessary for membrane integrity, photoprotection, orchestration of developmental programs, and anchoring biochemical functions to specific membrane systems. Such primary metabolites include, but are not limited to sterols, carotenoids,

chlorophyll, growth regulators, and the polyprenol substituents of dolichols, quinones, and proteins. Secondary metabolites mediate important interactions between plants and the environment, but are not necessary to the viability of the plant. Secondary metabolites include, but are not limited to tocopherols, monoterpenes, sesquiterpenes, and diterpenes.

For many years, it was accepted that IPP was synthesized through the well known acetate/mevalonate pathway. However, recent studies have demonstrated the occurrence of an alternative mevalonate-independent pathway for IPP biosynthesis (Horbach *et al.* (1993) *FEMS Microbiol. Lett.* 111:135-140; Rohmer et al., (1993) *Biochem J.* 295:517-524). This non-mevalonate pathway for IPP biosynthesis was initially characterized in bacteria and later also in green algae and higher plants (for recent reviews see Lichtenthaler et al.(1997) *Physiol. Plant.* 101:642-652 and Eisenreich et al.(1998) *Chem. Biol.* 5:R221-R233). The first reaction of the novel mevalonate-independent pathway is the condensation of (hvdroxyethyl)thiamin derived from pyruvate with the Cl aldehyde group of D-glyceraldehyde 3-phosphate to yield D-1-deoxyxylulose 5-phosphate (Broers (1994) *Ph.D. Thesis* Eidgenossische Technische Hochschule, Zurich, Switzerland; Rohmer et al., (1996) *J. Am. Chem. Soc.* 118:2564-2566). In Escherichia coli, D-1-deoxyxylulose (most likely in the form of D-1-deoxyxylulose 5-phosphate) is efficiently incorporated into the prenyl-side chain of menaquinone and ubiquinone (Broers, (1994) *supra*; Rosa Putra *et al.*, (1998) *Tetrahedron Lett.* 39:23-26). In plants, the incorporation of D-1-deoxyxylulose into isoprenoids has also been reported (Zeidler *et al.*, (1997) *Z. Naturforsch* 52c:15-23; Arigoni *et al.*, (1997) *Proc. Natl. Acad. Sci. USA* 94:10600-10605; Sagner *et al.*, (1998) *Chem. Commun.* 2:221-222). In addition, D-1-deoxyxylulose has also been described as a precursor for the biosynthesis of thiamin and pyridoxol. D-1-deoxyxylulose is the precursor molecule of the contiguous five-carbon unit (C4'-C4-C5-C5'-C5") of the thiazole ring of thiamin in E. coli (Therisod *et al.*, (1981)*Biochem. Biophys. Res. Comm.* 98:374-379; David *et al.* (1981) *J. Am. Chem. Soc.* 103:7341-7342) and in higher plant chloroplasts (Julliard and Douce, (1991) *Proc. Natl. Acad. Sci. USA* 88:2042-2045). The role of D-1-deoxyxylulose in the biosynthesis of pyridoxol in E. coli is also well documented (Hill *et al.*, (1989) *J. Am. Chem. Soc.* 111:1916-1917; Kennedy

et al., (1995) *J. Am. Chem. Soc.* 117:1661-1662; Hill *et al.*, (1996) *J. Biol. Chem.* 271:30426-30435). The cloning of genes encoding 1 -deoxy-D-xylulose 5-phosphate synthase has recently been reported in bacteria (Sprenger *et al.*, (1997) *Proc. Natl. Acad. Sci. USA* 94:12957-12962, Lois *et al.*, (1998) *Proc. Natl. Acad. Sci. USA* 95:2105-2110) and plants (Lange *et al.*, (1998) *Proc. Natl. Acad. Sci. USA* 95:2100-2104; Bouvier *et al.*, (1998) *Plant Physiol.* 117:1423-1431). Figure 2 provides a schematic representation of the isoprenoid pathways.

Although the intermediates between 1-deoxy-D-xylulose 5-phosphate and IPP have not yet been characterized, 2-C-methyl-D-erythriyol 4-phosphate has been proposed by Rohmer and co-workers as the first committed precursor for isoprenoid biosynthesis in bacteria (Duvold *et al.*, (1997) *Tetrahedron Lett.* 38:4769-4772; Duvold *et al.*, (1997) *Tetrahedron Lett.* 38:6181-6184). The enzyme 1-deoxy-D-xylulose 5-phosphate reductoisomerase, catalyzing the conversion of 1-D-deoxy-D-xylulose 5-phosphate into 2-C-methyl-D-erythhyol 4-phosphate, has been recently cloned and characterized in *E. coli* (Takahashi *et al.*, (1998) *Proc. Natl. Acad. Sci. USA* 95:99879-9884). The biosynthesis of 2-C-methyl-D-erythitol in plants by an intramolecular rearrangement of 1 -deoxy-D-xylulose 5-phosphate has recently been reported by Sagner *et al.* (1998) *Tetrahedron Lett.* 39:23-26 and Sagner *et al.* (1998) *Chem Commun.* 2:221-222.

The present invention provides polynucleotide and polypeptide sequences involved in the production of 2-C-Methyl-D-erythritol-4- phosphate from 1-deoxyxylulose-5-phosphate, referred to as 1-deoxy-D-xylulose 5-phosphate reductoisomerase or *dxr*. Also provided in the present invention are constructs and methods for the production of altered expression of *dxr* in host cells, as well as methods for the modification of the isoprenoid pathway, including modification of the biosynthetic flux through the isoprenoid pathway, and for the production of specific classes of isoprenoids in host cells.

Isolated Polynucleotides, Proteins, and Polypeptides

A first aspect of the present invention relates to isolated *dxr* polynucleotides. The polynucleotide sequences of the present invention include isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group of sequences set forth in the Sequence Listing and to other polynucleotide sequences closely related to such sequences and variants thereof.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence as set forth in the Sequence Listing. The invention also provides the coding sequence for the mature polypeptide or a fragment thereof, as well as the coding sequence for the mature polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, pro-, or prepro- protein sequence. The polynucleotide can also include non-coding sequences, including for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:

$$X\text{-}(R_1)_n\text{-}(R_2)\text{-}(R_3)_n\text{-}Y$$

wherein, at the 5' end, X is hydrogen, and at the 3' end, Y is hydrogen or a metal, R_1 and R_3 are any nucleic acid residue, n is an integer between 1 and 3000, preferably between 1 and 1000 and R_2 is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in the Sequence Listing and preferably those of SEQ ID NO:1. In the formula, R_2 is oriented so that its 5' end residue is at the left, bound to R_1, and its 3' end residue is at the right, bound to R_3. Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The invention also relates to variants of the polynucleotides described herein that encode for variants of the polypeptides of the invention. Variants that are fragments of

the polynucleotides of the invention can be used to synthesize full-length polynucleotides of the invention. Preferred embodiments are polynucleotides encoding polypeptide variants wherein 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues of a polypeptide sequence of the invention are substituted, added or deleted, in any combination. Particularly preferred are substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the polynucleotide or polypeptide.

Further preferred embodiments of the invention that are at least 50%, 60%, or 70% identical over their entire length to a polynucleotide encoding a polypeptide of the invention, and polynucleotides that are complementary to such polynucleotides. More preferable are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the invention and polynucleotides that are complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length are particularly preferred, those at least 95% identical are especially preferred. Further, those with at least 97% identity are highly preferred and those with at least 98% and 99% identity are particularly highly preferred, with those at least 99% being the most highly preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptides encoded by the polynucleotides set forth in the Sequence Listing.

The invention further relates to polynucleotides that hybridize to the above-described sequences. In particular, the invention relates to polynucleotides that hybridize under stringent conditions to the above-described polynucleotides. As used herein, the terms "stringent conditions" and "stringent hybridization conditions" mean that hybridization will generally occur if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at approximately 65°C. Other hybridization and wash conditions are well known and are

exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, cold Spring Harbor, NY (1989), particularly Chapter 11.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set for in the Sequence Listing under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers as described herein.

As discussed herein regarding polynucleotide assays of the invention, for example, polynucleotides of the invention can be used as a hybridization probe for RNA, cDNA, or genomic DNA to isolate full length cDNAs or genomic clones encoding a polypeptide and to isolate cDNA or genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in the Sequence Listing. Such probes will generally comprise at least 15 bases. Preferably such probes will have at least 30 bases and can have at least 50 bases. Particularly preferred probes will have between 30 bases and 50 bases, inclusive.

The coding region of each gene that comprises or is comprised by a polynucleotide sequence set forth in the Sequence Listing may be isolated by screening using a DNA sequence provided in the Sequence Listing to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to identify members of the library which hybridize to the probe. For example, synthetic oligonucleotides are prepared which correspond to the dxr sequences. The oligonucleotides are used as primers in polymerase chain reaction (PCR) techniques to obtain 5' and 3' terminal sequence of *dxr* genes. Alternatively, where oligonucleotides of low degeneracy can be prepared from particular dxr peptides, such probes may be used directly to screen gene libraries for dxr gene sequences. In particular, screening of cDNA libraries in phage vectors is useful in such methods due to lower levels of background hybridization.

Typically, a dxr sequence obtainable from the use of nucleic acid probes will show 60-70% sequence identity between the target dxr sequence and the encoding sequence used as a probe. However, lengthy sequences with as little as 50-60% sequence identity may also be obtained. The nucleic acid probes may be a lengthy fragment of the nucleic acid sequence, or may also be a shorter, oligonucleotide probe. When longer nucleic acid fragments are employed as probes (greater than about 100 bp), one may screen at lower stringencies in order to obtain sequences from the target sample which have 20-50% deviation (i.e., 50-80% sequence homology) from the sequences used as probe. Oligonucleotide probes can be considerably shorter than the entire nucleic acid sequence encoding a dxr enzyme, but should be at least about 10, preferably at least about 15, and more preferably at least about 20 nucleotides. A higher degree of sequence identity is desired when shorter regions are used as opposed to longer regions. It may thus be desirable to identify regions of highly conserved amino acid sequence to design oligonucleotide probes for detecting and recovering other related dxr genes. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (*See*, Gould, *et al.*, *PNAS USA* (1989) *86*:1934-1938.).

Another aspect of the present invention relates to dxr polypeptides. Such polypeptides include isolated polypeptides set forth in the Sequence Listing, as well as polypeptides and fragments thereof, particularly those polypeptides which exhibit dxr activity and also those polypeptides which have at least 50%, 60% or 70% identity, preferably at least 80% identity, more preferably at least 90% identity, and most preferably at least 95% identity to a polypeptide sequence selected from the group of sequences set forth in the Sequence Listing, and also include portions of such polypeptides, wherein such portion of the polypeptide preferably includes at least 30 amino acids and more preferably includes at least 50 amino acids.

"Identity", as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match

between strings of such sequences. "Identity" can be readily calculated by known methods including, but not limited to, those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M. and Griffin, H.G., eds., Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press (1987); *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., Stockton Press, New York (1991); and Carillo, H., and Lipman, D., SIAM *J Applied Math*, 48:1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available programs. Computer programs which can be used to determine identity between two sequences include, but are not limited to, GCG (Devereux, J., et al., *Nucleic Acids Research* 12(1):387 (1984); suite of five BLAST programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, *Trends in Biotechnology, 12:* 76-80 (1994); Birren, *et al., Genome Analysis, 1:* 543-559 (1997)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH, Bethesda, MD 20894; Altschul, S., *et al., J. Mol. Biol.*, 215:403-410 (1990)). The well known Smith Waterman algorithm can also be used to determine identity.

Parameters for polypeptide sequence comparison typically include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci USA* 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters along with no penalty for end gap are the default parameters for peptide comparisons.

Parameters for polynucleotide sequence comparison include the following:

Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970)

Comparison matrix: matches = +10; mismatches = 0

Gap Penalty: 50

Gap Length Penalty: 3

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters are the default parameters for nucleic acid comparisons.

The invention also includes polypeptides of the formula:

$$X\text{-}(R_1)_n\text{-}(R_2)\text{-}(R_3)_n\text{-}Y$$

wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, n is an integer between 1 and 1000, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in the Sequence Listing and preferably those encoded by the sequences provided in SEQ ID NO:2. In the formula, R_2 is oriented so that its amino terminal residue is at the left, bound to R_1, and its carboxy terminal residue is at the right, bound to R_3. Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

Polypeptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising a sequence selected from the group of a sequence contained in the Sequence Listing set forth herein .

The polypeptides of the present invention can be mature protein or can be part of a fusion protein.

Fragments and variants of the polypeptides are also considered to be a part of the invention. A fragment is a variant polypeptide which has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides. The fragments can be "free-standing" or comprised within a larger polypeptide of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically active fragments which are

those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that are antigenic or immunogenic in an animal, particularly a human.

Variants of the polypeptide also include polypeptides that vary from the sequences set forth in the Sequence Listing by conservative amino acid substitutions, substitution of a residue by another with like characteristics. In general, such substitutions are among Ala, Val, Leu and Ile; between Ser and Thr; between Asp and Glu; between Asn and Gln; between Lys and Arg; or between Phe and Tyr. Particularly preferred are variants in which 5 to 10; 1 to 5; 1 to 3 or one amino acid(s) are substituted, deleted, or added, in any combination.

Variants that are fragments of the polypeptides of the invention can be used to produce the corresponding full length polypeptide by peptide synthesis. Therefore, these variants can be used as intermediates for producing the full-length polypeptides of the invention.

The polynucleotides and polypeptides of the invention can be used, for example, in the transformation of host cells, such as plant host cells, as further discussed herein.

The invention also provides polynucleotides that encode a polypeptide that is a mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one polypeptide chain). Such sequences can, for example, play a role in the processing of a protein from a precursor to a mature form, allow protein transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production. It is contemplated that cellular enzymes can be used to remove any additional amino acids from the mature protein.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. The inactive precursors generally are activated when the prosequences are removed. Some or all of the prosequences may be removed prior to activation. Such precursor protein are generally called proproteins.

Constructs and Methods of Use

Of particular interest is the use of the nucleotide sequences in recombinant DNA constructs to direct the transcription or transcription and translation (expression) of the dxr sequences of the present invention in a host cell. The expression constructs generally comprise a promoter functional in a host cell operably linked to a nucleic acid sequence encoding a dxr of the present invention and a transcriptional termination region functional in a host cell. Host cells of particular interest in the present invention include, but are not limited to, fungal cells, yeast cells, bacterial cells, mammalian cells, and plant cells.

A first nucleic acid sequence is "operably linked" or "operably associated" with a second nucleic acid sequence when the sequences are so arranged that the first nucleic acid sequence affects the function of the second nucleic-acid sequence. Preferably, the two sequences are part of a single contiguous nucleic acid molecule and more preferably are adjacent. For example, a promoter is operably linked to a gene if the promoter regulates or mediates transcription of the gene in a cell.

Those skilled in the art will recognize that there are a number of promoters which are functional in plant cells, and have been described in the literature. Chloroplast and plastid specific promoters, chloroplast or plastid functional promoters, and chloroplast or plastid operable promoters are also envisioned.

One set of plant functional promoters are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant organs. Enhanced or duplicated versions of the CaMV35S and FMV35S promoters are useful in the practice of this invention (Odell, *et al.* (1985) *Nature* 313:810-812; Rogers, U.S. Patent Number 5,378, 619). In addition, it may also be preferred to bring about expression of the dxr gene in specific tissues of the plant, such as leaf, stem, root, tuber, seed, fruit, etc., and the promoter chosen should have the desired tissue and developmental specificity.

Of particular interest is the expression of the nucleic acid sequences of the present invention from transcription initiation regions which are preferentially expressed in a

plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl *et al.*, *Seed Sci. Res. 1*:209:219 (1991)), phaseolin, zein, soybean trypsin inhibitor, ACP, stearoyl-ACP desaturase, soybean α' subunit of β-conglycinin (soy 7s, (Chen *et al.*, *Proc. Natl. Acad. Sci.*, 83:8560-8564 (1986))) and oleosin.

It may be advantageous to direct the localization of proteins conferring dxr to a particular subcellular compartment, for example, to the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, where the genes of interest of the present invention will be targeted to plastids, such as chloroplasts, for expression, the constructs will also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, where the gene of interest is not directly inserted into the plastid, the expression construct will additionally contain a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.* (1991) *Plant Mol. Biol. Rep. 9*:104-126; Clark *et al.* (1989) *J. Biol. Chem. 264*:17544-17550; della-Cioppa *et al.* (1987) *Plant Physiol. 84*:965-968; Romer *et al.* (1993) *Biochem. Biophys. Res Commun. 196*:1414-1421; and, Shah *et al.* (1986) *Science 233*:478-481.

Depending upon the intended use, the constructs may contain the nucleic acid sequence which encodes the entire dxr protein, or a portion thereof. For example, where antisense inhibition of a given dxr protein is desired, the entire dxr sequence is not required. Furthermore, where dxr sequences used in constructs are intended for use as probes, it may be advantageous to prepare constructs containing only a particular portion of a dxr encoding sequence, for example a sequence which is discovered to encode a highly conserved dxr region.

The skilled artisan will recognize that there are various methods for the inhibition of expression of endogenous sequences in a host cell. Such methods include, but are not limited to, antisense suppression (Smith, *et al.* (1988) *Nature* 334:724-726) , co-suppression (Napoli, *et al.* (1989) *Plant Cell* 2:279-289), ribozymes (PCT Publication WO 97/10328), and combinations of sense and antisense Waterhouse, *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964. Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence.

Regulatory transcript termination regions may be provided in plant expression constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the dxr or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. The skilled artisan will recognize that any convenient transcript termination region which is capable of terminating transcription in a plant cell may be employed in the constructs of the present invention.

Alternatively, constructs may be prepared to direct the expression of the dxr sequences directly from the host plant cell plastid. Such constructs and methods are known in the art and are generally described, for example, in Svab, *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530 and Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917 and in U.S. Patent Number 5,693,507.

The constructs of the present invention can also be used in methods for altering the flux through the isoprenoid pathway with additional constructs for the expression of additional genes involved in the production of isoprenoids. Such sequences include, but are not limited to 1-deoxyxylulose 5-phosphate synthase.

Furthermore, the constructs of the present invention can be used in transformation methods with additional constructs providing for the expression of additional nucleic acid sequences encoding proteins in the production of specific isoprenoids, such as tocopherols, carotenoids, sterols, monoterpenes, sesquiterpenes, and diterpenes. Nucleic

acid sequences involved in the production of carotenoids and methods are described for example in PCT publication WO 99/07867. Nucleic acid sequences involved in the production of tocopherols include, but are not limited to gamma-tocpherol methyltransferase (Shintani, *et al.* (1998) *Science* 282(5396):2098-2100), tocopherol cyclase, and tocopherol methyltransferase, phytyl prenyltransferase, geranylgeranylpyrophosphate hydrogenase, geranylgeranylpyrophosphate synthase.

A plant cell, tissue, organ, or plant into which the recombinant DNA constructs containing the expression constructs have been introduced is considered transformed, transfected, or transgenic. A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a dxr nucleic acid sequence.

Plant expression or transcription constructs having a dxr encoding sequence as the DNA sequence of interest for increased or decreased expression thereof may be employed with a wide variety of plant life. Particularly preferred plants for use in the methods of the present invention include, but are not limited to: *Acacia,* alfalfa, aneth, apple, apricot, artichoke, arugula, asparagus, avocado, banana, barley, beans, beet, blackberry, blueberry, broccoli, brussels sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, celery, cherry, chicory, cilantro, citrus, clementines, coffee, corn, cotton, cucumber, Douglas fir, eggplant, endive, escarole, eucalyptus, fennel, figs, garlic, gourd, grape, grapefruit, honey dew, jicama, kiwifruit, lettuce, leeks, lemon, lime, Loblolly pine, mango, melon, mushroom, nectarine, nut, oat, oil palm, oil seed rape, okra, onion, orange, an ornamental plant, papaya, parsley, pea, peach, peanut, pear, pepper, persimmon, pine, pineapple, plantain, plum, pomegranate, poplar, potato, pumpkin, quince, radiata pine, radicchio, radish, raspberry, rice, rye, sorghum, Southern pine, soybean, spinach, squash, strawberry, sugarbeet, sugarcane, sunflower, sweet potato, sweetgum, tangerine, tea, tobacco, tomato, triticale, turf, turnip, a vine, watermelon, wheat, yams, and zucchini. Particularly preferred are plants involved in the production of vegetable oils for edible and industrial uses. Most especially preferred are temperate oilseed crops. Temperate oilseed crops of interest include, but are not limited to, rapeseed (Canola and High Erucic

Acid varieties), sunflower, safflower, cotton, soybean, peanut, coconut and oil palms, and corn. Depending on the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledyons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

Of particular interest, is the use of dxr constructs in plants to produce plants or plant parts, including, but not limited to leaves, stems, roots, reproductive, and seed, with a modified content of tocopherols in plant parts having transformed plant cells.

For immunological screening, antibodies to the protein can be prepared by injecting rabbits or mice with the purified protein or portion thereof, such methods of preparing antibodies being well known to those in the art. Either monoclonal or polyclonal antibodies can be produced, although typically polyclonal antibodies are more useful for gene isolation. Western analysis may be conducted to determine that a related protein is present in a crude extract of the desired plant species, as determined by cross-reaction with the antibodies to the encoded proteins. When cross-reactivity is observed, genes encoding the related proteins are isolated by screening expression libraries representing the desired plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

To confirm the activity and specificity of the proteins encoded by the identified nucleic acid sequences as dxr enzymes, *in vitro* assays are performed in insect cell cultures using baculovirus expression systems. Such baculovirus expression systems are known in the art and are described by Lee, *et al.* U.S. Patent Number 5,348,886, the entirety of which is herein incorporated by reference.

In addition, other expression constructs may be prepared to assay for protein activity utilizing different expression systems. Such expression constructs are transformed into yeast or prokaryotic host and assayed for dxr activity. Such expression systems are known in the art and are readily available through commercial sources.

In addition to the sequences described in the present invention, DNA coding sequences useful in the present invention can be derived from algae, fungi, bacteria, mammalian sources, plants, etc. Homology searches in existing databases using signature sequences corresponding to conserved nucleotide and amino acid sequences of dxr can be employed to isolate equivalent, related genes from other sources such as plants and microorganisms. Searches in EST databases can also be employed. Furthermore, the use of DNA sequences encoding enzymes functionally enzymatically equivalent to those disclosed herein, wherein such DNA sequences are degenerate equivalents of the nucleic acid sequences disclosed herein in accordance with the degeneracy of the genetic code, is also encompassed by the present invention. Demonstration of the functionality of coding sequences identified by any of these methods can be carried out by complementation of mutants of appropriate organisms, such as *Synechocystis, Shewanella, yeast, Pseudomonas, Rhodobacteria*, etc., that lack specific biochemical reactions, or that have been mutated. The sequences of the DNA coding regions can be optimized by gene resynthesis, based on codon usage, for maximum expression in particular hosts.

The method of transformation in obtaining such transgenic plants is not critical to the instant invention, and various methods of plant transformation are currently available. Furthermore, as newer methods become available to transform crops, they may also be directly applied hereunder. For example, many plant species naturally susceptible to *Agrobacterium* infection may be successfully transformed via tripartite or binary vector methods of *Agrobacterium* mediated transformation. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-DNA borders may find use with other modes of transformation. In addition, techniques of microinjection, DNA particle bombardment, and electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) will be inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, there being broad host range vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta, *et al.*, (*Proc. Nat. Acad. Sci., U.S.A.* (1980) *77*:7347-7351) and EPA 0 120 515, which are incorporated herein by reference. Alternatively, one may insert the sequences to be expressed in plant cells into a vector containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride, *et al.* (*Plant Mol. Biol.* (1990) *14*:269-276), wherein the pRiHRI (Jouanin, *et al., Mol. Gen. Genet.* (1985) 201:370-374) origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed Agrobacterium and transformed plant cells. A number of markers have been developed for use with plant cells, such as

resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

There are several possible ways to obtain the plant cells of this invention which contain multiple expression constructs. Any means for producing a plant comprising a construct having a DNA sequence encoding the expression construct of the present invention, and at least one other construct having another DNA sequence encoding an enzyme are encompassed by the present invention. For example, the expression construct can be used to transform a plant at the same time as the second construct either by inclusion of both expression constructs in a single transformation vector or by using separate vectors, each of which express desired genes. The second construct can be introduced into a plant which has already been transformed with the dxr expression construct, or alternatively, transformed plants, one expressing the dxr construct and one expressing the second construct, can be crossed to bring the constructs together in the same plant.

The nucleic acid sequences of the present invention can be used in constructs to provide for the expression of the sequence in a variety of host cells, both prokaryotic eukaryotic. Host cells of the present invention preferably include monocotyledenous and dicotyledenous plant cells.

In general, the skilled artisan is familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of recombinant organisms and the screening and isolating of clones, (see for example,

Sambrook *et al., Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Press (1989); Maliga *et al., Methods in Plant Molecular Biology,* Cold Spring Harbor Press (1995), the entirety of which is herein incorporated by reference; Birren *et al., Genome Analysis: Analyzing DNA,* 1, Cold Spring Harbor, New York, the entirety of which is herein incorporated by reference).

Methods for the expression of sequences in insect host cells are known in the art. Baculovirus expression vectors are recombinant insect viruses in which the coding sequence for a chosen foreign gene has been inserted behind a baculovirus promoter in place of the viral gene, e.g., polyhedrin (Smith and Summers, U.S. Pat. No., 4,745,051, the entirety of which is incorporated herein by reference). Baculovirus expression vectors are known in the art, and are described for example in Doerfler, *Curr. Top. Microbiol. Immunol. 131*:51-68 (1968); Luckow and Summers, *Bio/Technology 6*:47-55 (1988a); Miller, *Annual Review of Microbiol. 42*:177-199 (1988); Summers, *Curr. Comm. Molecular Biology*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1988); Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures,* Texas Ag. Exper. Station Bulletin No. 1555 (1988), the entireties of which is herein incorporated by reference)

Methods for the expression of a nucleic acid sequence of interest in a fungal host cell are known in the art. The fungal host cell may, for example, be a yeast cell or a filamentous fungal cell. Methods for the expression of DNA sequences of interest in yeast cells are generally described in "Guide to yeast genetics and molecular biology", Guthrie and Fink, eds. Methods in enzymology , Academic Press, Inc. Vol 194 (1991) and Gene expression technology", Goeddel ed, Methods in Enzymology, Academic Press, Inc., Vol 185 (1991).

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC, Manassas, VA), such as HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include, but are not limited to, viral promoters such as that from Simian Virus 40 (SV40) (Fiers *et al., Nature*

273:113 (1978), the entirety of which is herein incorporated by reference), Rous sarcoma virus (RSV), adenovirus (ADV) and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly-A addition sequences. Enhancer sequences which increase expression may also be included and sequences which promote amplification of the gene may also be desirable (for example methotrexate resistance genes).

Vectors suitable for replication in mammalian cells are well known in the art, and may include viral replicons, or sequences which insure integration of the appropriate sequences encoding epitopes into the host genome. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (*see*, for example, Mackett *et al, J Virol. 49*:857 (1984); Chakrabarti *et al., Mol. Cell. Biol. 5*:3403 (1985); Moss, In: *Gene Transfer Vectors For Mammalian Cells* (Miller and Calos, eds., Cold Spring Harbor Laboratory, N.Y., p. 10, (1987); all of which are herein incorporated by reference in their entirety).

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

EXAMPLE 1: Synthesis of 2-C-methyl-D-erythritol

2-C-Methyl-D-erythritol with a ca 80% e.e. was synthesized according to a Duvold, *et al.* (1997) *Tetrahedron Lett* 38:4769-4772 and Duvold, *et al.* (1997) *Tetrahedron Lett* 38:6181-6184) adapted to the production of larger amounts. A solution of 3-methyl-2(5*H*)-furanone (200 mg, 2 mmol) in dry ether (20 ml) was added at 0°C over a period of 15 min to a stirred suspension of $LiAlH_4$ (46 mg, 1.2 mmol) in dry ether (20 ml) under argon. The reaction mixture was stirred at O°C for further 2 h. A saturated solution of NH_4Cl (2 ml) was slowly added until the excess of $LiAlH_4$ was destroyed.

After acidification with a 1M HCl solution until all aluminum salts were dissolved, the aqueous phase was extracted with ethyl acetate (6 x 20 ml). The combined organic layers were washed with saturated brine and dried over anhydrous Na_2SO_4. After removal of the solvent under reduced pressure, the crude diol (1 77 mg) dissolved in methylene chloride (20 ml) was directly acetylated for 15 min with a mixture of acetic anhydride/triethylamine (2:3, v/v, 1 ml) in presence of catalytic amounts of dimethylaminopyridine (12 mg). Solvent and excess or reagents were evaporated under reduced pressure. Flash column chromatography (Still et al., 1978) (hexane/ethyl acetate, 4:1, v/v) afforded pure diacetate (330 mg, 86 %). Enantioselective dihydroxylation of diacetate (300 mg, 1.6 mmol) was performed by stirring at O°C in ter-butanol/water (1:1, v/v, 6 ml) in the presence of the chiral osmylation reagent AD-mix-b (2.5 g) and $CH_3SO_2NH_2$ (152 mg, 1.6 mmol). After 24 hours, the reaction was quenched with solid $Na2SO_3$ and additional stirring for 30 minutes. Repeated extraction with ethyl acetate (6 x 20 ml) and flash chromatography (ethyl acetate) afforded a mixture only containing 2-C-methyl-D-erythritol diacetates (resulting from partial intramolecular transesterifications) (312 mg, 88 % yield). Quantitative deacetylation was performed overnight at room temperature in the presence of basic Amberlyst A-26 (OH- form) (150 mg for 1 mmol) in methanol (30 ml) (Reed et al., 1981) Filtration of the resin and evaporation of the solvent directly afforded pure 2-C-methyl-D-erythritol (1 90 mg, 75 % overall yield).

EXAMPLE 2: Site-Directed Marker Insertion Mutagenesis of the *dxr* gene of *E. coli*

The region extending from the 5'-region of the*dxr* gene to the 3'-flanking region of the *yae*S gene was amplified by PCR using genomic DNA isolated from the wild type *E. coli* strain W3110 (Kohara et al., 1987) and the primers P1(5'-CTCTGGATGT CATATGAAGCAACTC-3' (SEQ ID NO:3); the underlined ATG corresponds to the translation start codon of the *dxr* gene) and P2 (5'-CCGCATAACACCGCCAACC-3' (SEQ ID NO:4); located at the 3'-flanking region of the *yae*S gene). The reaction mixture

for the PCR was prepared in a final volume of 50 µl, containing the DNA template (100 ng), 0.5 µM of each primer, 200 µM of each deoxynucleoside triphosphate, 20 mM of Tris-HCl adjusted to pH 8.8, 2 mM Of $MgSO_4$, 10 mM of KCI, 10 mM of $(NH_4)_2SO_4$, 0.1 mg/ml of BSA and 0.1% Triton X-100. The sample was covered with mineral oil, incubated at 94°C for 3 min and cooled to 8O°C. *Pfu* DNA polymerase (1.25 units, Stratagene) was added and the reaction mixture was incubated for 30 cycles consisting of 45 sec at 94°C, 45 sec at 59°C and 10 min at 72°C, followed by a final step of 10 min at 72°C. After amplification, adenines were added to the 3' ends of the PCR product as indicated by the manufacturers protocol and the adenylated product was cloned into the pGEM-T vector (Promega), to create plasmid pMJ1. The CAT (chloramphenicol acetyl transferase) gene present in plasmid pCAT19 (Fuqua, 1992) was excised by digestion with *Pst*I and *Xba*I, treated with T4 DNA polymerase and cloned into the unique *Asu*II site present in the *dxr* gene by blunt end ligation (after treatment with T4 DNA polymerase), resulting plasmid pMJ2. Restriction enzyme mapping was used to identify the clones in which the CAT gene was in the same orientation than the *dxr* gene. Plasmid pMJ3 was constructed by subcloning the *Spe*I-*Sph*I fragment excised from plasmid pMJ2 into the *Nhe*I-*Sph*I sites of plasmid pBR322. Plasmid pMJ3 was linearized by digestion with *Pst*I, incubated with calf intestinal alkaline phosphatase (GibcoBRL) and purified by agarose gel electrophoresis. Two µg of the purified linear plasmid pMJ3 DNA were used to transform E coli strain JC7623 (Winans et al., 1985). Transformed cells were plated onto LB plates (Ausubel et al. 1987) supplemented with 2 mM of 2-C-methyl-D-erythritol (ME) and chloramphenicol (17 µg/mL). Colonies showing both chloramphenicol resistance and ME auxotrophy were selected for further studies. The presence of the CAT gene insertion into the *dxr* gene was checked by PCR using primers P3 (5'-GCACACTTCCACTGTGTGTG-3' (SEQ ID NO:5), located at the 5'-region of the *frr* gene) and P2. One of these colonies, designated as strain JC7623*dxr*:CAT, was used for the complementation studies.

EXAMPLE 3: Rapid Amplification of cDNA Ends (RACE)

To identify putative plant nucleic acid sequences encoding homologues of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), the Non-Redundant database of the National Center for Biotechnology Information (NCBI) was searched with the TBLASTN program, using the complete amino acid sequence of the recently cloned DXR from *Escherichia coli* (Takahashi et al., 1998) as a query. A significant level of identity (40-64%) was found between this query and the amino acid sequence encoded by seven predicted exons of the *A. thaliana* genomic clone MQB2 (Accession number ABOO9053).

To confirm the existence of mRNA sequences corresponding to the putative *A. thaliana* DXR gene, the EST database of the NCBI (dbEST) was searched with the BLASTN program using as a query the nucleotide sequence of clone MOB2 extending from nucleotides 29247 to 31317. Two A. thaliana EST clones (12OE8T7 and 65FllXP3', accession numbers T43949 and AA586087,respectively) containing nucleotide sequences identical to different regions of the query were found. Sequencing of the cDNA inserts revealed the two clones were overlapping. The longest cDNA contained an open reading frame encoding a polypeptide of 329 residues showing an identity of 41.6% (similar of 53.2%) with the C-terminal region of the *E. coli* DXR, thus indicating that the two cDNAs encoded truncated versions of the putative *A. thaliana* enzyme.

Total RNA from l2-days-old light-grown *Arabidopsis thaliana* (var. Columbia) seedlings was purified as described (Dean et al., 1985). Rapid amplification of cDNA ends (RACE) was carried out with the 5'-RACE-System (Version 2.0) from Life Technologies/Gibco BRL, following the instructions of the supplier. The first strand of cDNA was synthesized using 1 μg of the RNA sample as template and the oligonucleotide DXR-GSP1 (5'-ATTCGAACCAGCAGCTAGAG-3' (SEQ ID NO:6), complementary to nucleotides +767 to +786 of the sequence shown in SEQ ID NO:1 as specific downstream primer. After purification and homopolymeric tailing of the cDNA, two nested PCR reactions were performed. In the first PCR, the specific downstream primer was the oligonucleotide DXR-GSP2 (5'-CCAGTAGATCCAACGATAGAG-3' (SEQ ID NO:7), complementary to nucleotides +530 to +550 of the sequence shown in

SEQ ID NO:1) and the upstream primer was the oligonucleotide 5'-RACE-AAP (supplied in the kit). In the second PCR, the specific downstream primer was the oligonucleotide DXR-GSP3 (5'-GGCCATGCTGGAGGAGGTTG-3' (SEQ ID NO:8), complementary to nucleotides +456 to +475 of the sequence shown in SEQ ID NO:1) and the upstream primer was the oligonucleotide AUAP (supplied in the kit). In both PCR reactions the amplification process was initiated by denaturation of the sample (3 min at 94°C), cooling to 80°C and addition of *Taq* DNA polymerase. The reaction mixture of the first PCR was incubated for 15 cycles consisting of 30 sec at 94°C, 30 sec at 55°C and 1 min at 72°C, followed by a final step of 5 min at 72°C. The sample obtained was diluted one to ten in the reaction mixture of the second PCR and incubated for 30 cycles consisting of 30 sec at 94°C, 30 sec at 61°C and 1 min at 72°C, with a final step of 5 min at 72°C. The final amplification products were purified by agarose gel electrophoresis, cloned into plasmid pBluescript SK+ and sequenced (SEQ ID NO:1).

EXAMPLE 4: Cloning of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase cDNA from *Arabidopsis thaliana*

To define the 5'-region of the putative DXR gene, the corresponding transcription start site was mapped by using the RACE technique. Primers were designed on the basis of the alignment between the DXR from E. coli and the amino acid sequence deduced from the A. thaliana genomic clone. The deduced amino acid sequence from the *Arabidopsis* dxr nucleic acid sequence (SEQ ID NO:1) is provided in SEQ ID NO:2. The first strand of cDNA was synthesized using RNA from A. thaliana seedlings as a template and the oligonucleotide DXR-GSP1 as primer. This oligonucleotide was complementary to the region between positions +767 and +786 of the genomic sequence shown in SEQ ID NO:1. Subsequently, two nested PCR reactions were carried out to ampl4 the 5' end of the mRNA. The downstream specific primers used for the first and second nested PCR reactions were complementary to the regions extending from positions +530 to +550 (primer DXR-GSP2) and +456 to +475 (primer DXR-GSP3), respectively. Four clones

corresponding to the major amplification product were sequenced and found to have the same 5'-end, which corresponds to the adenine at position +1 in the genomic sequence shown in SEQ ID NO:1.

A cDNA containing the whole coding sequence of the *Arabidopsis* DXR was amplified by two consecutive PCR reactions from a cDNA library derived from the *A. thaliana* (var. Columbia) cell suspension line T87. An aliquot of the library was ethanol-precipitated and resuspended in water. The reaction mixture for the first PCR was prepared in a final volume of 25 µl containing the DNA template (equivalent to $4x10^5$ pfu of cDNA library), 0.5 µM of the upstream primer DXR-34 (5'-CAAGAGTAGTAGTGCGGTTCTCTGG-3' (SEQ ID NO:9), corresponding to nucleotides +34 to +58 of the sequence shown in SEQ ID NO:1), 0.5 ;µM of the downstream primer DXR-E2 (5'-CAGTTTGGCTTGTTCGGATCACAG-3' (SEQ ID NO:10), complementary to nucleotides +3146 to + 3169 of the sequence shown in SEQ ID NO:1), 200 µM of each deoxynucleoside triphosphate, 20 mM of Tris-HCI adjusted to pH 8.8, 2 mM Of $MgSO_4$, 10 mM of KCI, 10 MM of $(NH_4)_2SO_4$, 0.1 mg/ml of BSA and 0.1 % Triton X-100. The sample was covered with mineral oil, incubated at 94°C for 3 min and cooled to 80°C. Pfu DNA polymerase (1.25 units, Stratagene) was added and the reaction mixture was incubated for 35 cycles consisting of 30 sec at 94°C, 40 sec at 55°C and 6.5 min at 72°C, followed by a final step of 15 min at 72°C. The reaction mixture was diluted one to ten with water and 5 µl were used as a template for the second PCR that was performed using the same conditions as described for the previous amplification, except that the volume of the reaction mixture was increased to 50 µl and the number of cycles was reduced to 15. The amplification product was purified by agarose gel electrophoresis and cloned into plasmid pBluescript SK+. The resulting plasmid was named pDXR-At.

Thus, a cDNA clone encoding the entire *A. thaliana* DXR was obtained by PCR from a cDNA library using primers DXR-34 and DXR-E2 corresponding to the regions extending from positions +34 to +58 and +3146 to +3169 of the genomic sequence, respectively. The identity of the amplified cDNA was confirmed by DNA sequencing, The alignment of the cDNA and the genomic sequences showed that the *A. thaliana* DXR

gene contains 12 exons and 11 introns which extend over a region of 3.2 Kb (SEQ ID NO:1).

The cloned cDNA encodes a protein of 477 amino acid residues with a predicted molecular mass of 52 kDa. The alignment of *A. thaliana* and *E. coli* DXR (Figure 1) reveals that the plant enzyme has a N-terminal extension of 79 residues with the typical features of plastid transit peptides (von Heijne et al., 1989). The two proteins show an identity of 42.7% (similarity of 54.3%).

EXAMPLE 5: Expression Construct Preparation

To express the *A. thaliana* DXR in *E. coli*, the region of the DXR cDNA encoding amino acid residues 81 to 477 was amplified by PCR from plasmid pDXR-At and cloned into a modified version of plasmid pBAD-GFPuv (Clontech). In this plasmid, expression is driven by the P_{BAD}promoter which can be induced with arabinose and repressed with glucose. First, plasmid pBAD-GFPuv was modified by removing the *Nde*I site located between pBR322ori and the araC coding region (position 4926-4931) by site-directed mutagenesis following the method of Kunkel et al. (Kunkel et al., 1987). The oligonucleotide pBAD-mut1 (5'-CTGAGAGTGCACCATCTGCGGTGTGAAATACC-3' (SEQ ID NO:11)) was used as mutagenic primer. The resulting plasmid was designated pBAD-Mi. Next, *Nde*I and *Eco*RI restriction sites were introduced at appropriate positions of the *A. thaliana* DXR cDNA by PCR, using the plasmid pDXR-At as template and the oligonucleotides 5'-MVKPI (5'-GGCATATGGTGAAACCCATCTCTATCGTTGGATC-3' (SEQ ID NO:12), complementary to nucleotides +522 to +544 of the sequence shown in SEQ ID NO:1; the underlined sequence contains the *Nde*I site) and DXR-END(5'-ACGAATTCATTATGCATGAACTGGCCTAGCACC-3' (SEQ ID NO:13), complementary to nucleotides+2997 to +3018 of the sequence shown in SEQ ID NO:1; the underlined sequence contains the *Eco*RI site) as mutagenic primers. The PCR amplification product was digested with *Nde*I and *Eco*RI and cloned into plasmid pBAD-

MI digested with the same restriction enzyme. This resulted in the substitution of the GFPuv coding sequence in plasmid pBAD-MI by the corresponding coding sequence of the. *A. thaliana* DXR. The resulting plasmid, designated pBAD-DXR, was introduced into strain XL1-Blue. Plasmid pBAD-MI, encoding GFPuv, was used as a control in the complementation studies.

EXAMPLE 6: Analysis of the *Arabidopsis thaliana dxr*

The function of the cloned *A. thaliana* DXR has been established by complementation analysis of an *E. coli* strain carrying a disruption in the dxr gene (strain JC7623dxr.:CAT) (see Example 2). This strain requires 2-C-methyl-D-erythritol (ME) for growth. For the complementation studies we used the region of the *A. thaliana* DXR extending from amino acids 81 to 477 of SEQ ID NO:2, which does not include the putative plastid transit peptide. The appropriate cDNA fragment was cloned into a derivative of plasmid pBAD-GFPuv, under the control of the PBAD promoter, and the resulting plasmid (pBAD-DXR) introduced into the JC7623dxr.-CAT strain. Expression from the PBAD promoter is inducible by arabinose and repressed by glucose. Induction with arabinose allows growth of strain JC7623dxr.-CAT harbouring plasmid PBAD-DXR in the absence of ME, whereas no growth was observed in the presence of glucose. Strain JC7623dxr.,:CAT carrying the control plasmid pBAD-MI does not grow in the presence of arabinose on medium lacking ME. Strain JC7623dxr.-CAT carrying either plasmid PBAD-DXR or pBAD-GFPuv grows on medium containing ME. These results unequivocally demonstrate that the cloned *A. thaliana* cDNA encodes a functional DXR.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same

extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.